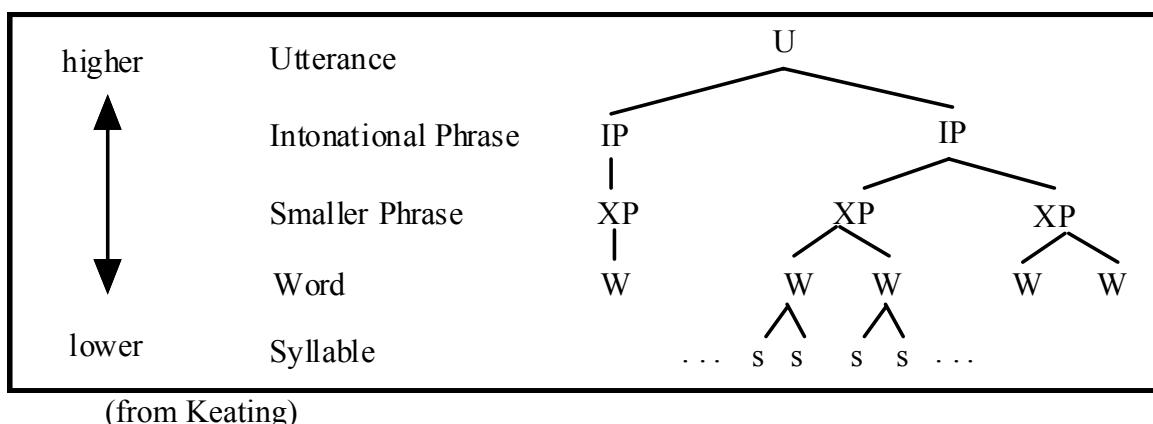


Prosody and Pitch

1. From a phonetic point of view, we observe that human speech cannot be fully characterized as the manifestation of sequences of phonemes, syllables, or words.
 - Pitch moves up and down in a non-random way, providing speech with recognizable melodical properties.
 - Segments or syllables may be shortened or lengthened, apparently in accordance with some underlying pattern.
 - Some syllables or words sound more prominent than others.
 - The stream of words is subdivided into phrases made up of words that seem to belong together. And phrases can sound as if they relate to one another, or, alternatively, as if they have nothing to do with each other.
2. Those properties of speech that cannot be derived from the underlying sequence of segments are called suprasegmental properties. They usually involve two or more segments and occur simultaneously with those segments.
 - e.g., F0, duration, amplitude
 - stress, length, tone, intonation
3. Prosody is a hierarchical organization of speech which is cued by suprasegmental acoustic features.

Prosody has both phonetic and phonological aspects.

- The phonological aspect is the hierarchical organization of segments into constituents with a pattern of relative prominences within these constituents.
Constituents include Intonational Phrase (IP), Prosodic Phrases (intermediate phrase, phonological phrase, accentual phrase, etc.), Prosodic Word, Foot, Syllable.



Phrase-level prominences include nuclear pitch accent > pre-nuclear pitch accent.
Word-level prominence distinguishes stressed syllables from unstressed syllables.

- The phonetic aspect is the set of acoustic parameters that provide evidence for prosodic organization.
 - Suprasegmental acoustic properties define suprasegmental features: stress, length, tone, intonation.
 - Stress is cued by duration and/or amplitude at the word level.
Stressed syllables are longer and louder.
 - Stress is cued by F0 at the phrase level (i.e., pitch accents).
Stressed words have pitch accents (usually high, sometimes low).
 - Intonation is largely F0, but duration and/or intensity may also be cues.
 - (Tone is cued by F0.)

There are also other phonetic reflexes of the organization.

- e.g., Pre-boundary duration is correlated with prosodic structure: pre-boundary lengthening increases for final boundaries at increasingly higher prosodic constituents.

Some such reflexes are realized segmentally (e.g., certain phonetic properties of individual segments may depend on the segment's prosodic position).

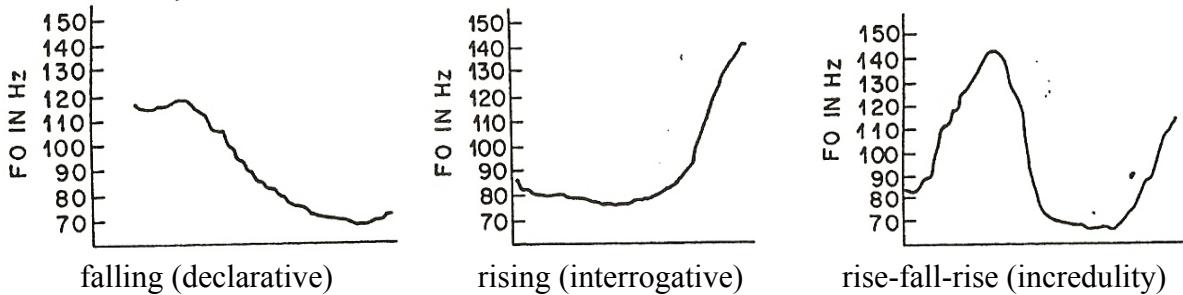
- Initial articulations for every constituent are stronger than medial ones, increasingly so in higher prosodic constituents. → domain-initial strengthening

Strengthening has lots of (segmental) acoustic correlates, e.g., increased closure duration, increased VOT (in Korean, but not French or Taiwanese)

The Linguistic Function of Pitch (F0)

4. Fundamental frequency – F0 – is the rate at which the vocal folds vibrate.
 - Smaller vocal folds vibrate faster → higher F0
 - Increased airflow → increase in pitch
 - An individual can raise F0 by pulling the vocal folds tighter so they vibrate faster.
5. F0 conveys paralinguistic information.
 - speaker's sex, age
 - speaker's emotional state
6. F0 may be affected by various linguistic factors.
 - F0 can differ by language.
 - e.g., average Japanese F0 is higher than English
 - F0 can vary according to the segment (these are not strictly suprasegmental properties).
 - e.g., high vowels → higher F0
 - e.g., consonant may have a particular F0
voiceless C (/p/, /s/) cause local raising of pitch
 - e.g., consonant may affect F0 of following V
in Yoruba, /g/ rises into V and /k/ falls into V

7. Intonation is the controlled modulation of voice pitch across a phrase or phrases.
 All languages seem to use at least some intonation to mark prosodic (syntactic?) information (i.e., it marks phrase boundaries). Intonation can signal information structure (i.e., what's new or old, what's important), differences in meaning (e.g., questions vs. statements), or non-linguistic information like attitude or emotion.
8. A first approach to description would be simply to show/describe abstract contours:



However, such forms of descriptions don't give us much means to describe more complex patterns, and they don't allow us to see layers of patterns (that reflect layers of linguistic structure).

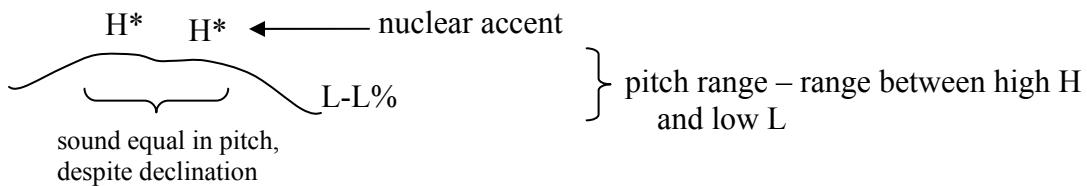
9. Intonational contours can also be described as a string of categorically distinct tonal elements – autosegmental-metrical models of intonation. In other words, a contour is a grammatically-governed concatenation of pitch targets.

English intonation is determined by three parameters: (1) pitch accent types, (2) pitch accent locations, and (3) phrasing. These parameters are independent of one another. (Beckman & Pierrehumbert, 1986)

- pitch accent (*): a tone (sequence) that aligns with a stressed syllable within a phrase. The syllable associated with the pitch accent is called an accented syllable. For a bitonal accent, the starred tone is associated with a stressed syllable.
 - English has two simple pitch accents: H*, L* and four bitonal accents: L+H*, L*+H, H*+L, H+L*.
 - A nuclear pitch accent is the last pitch accent in the intermediate phrase and is the most prominent syllable of the phrase.
- phrasal tones: tones associated with the edge of a phrase (i.e., not linked to a specific syllable)
 - phrase accent (-): H- or L-. Phrase accents belong to an intermediate phrase and are realized on syllables after the nuclear pitch accented word up to the end of the intermediate phrase.
 - boundary tone (%): H% or L%. Boundary tones belong to an intonational phrase and are realized on the very last syllable of the IP. (IP boundaries are also marked by phrase final lengthening and may be followed by a pause.)

10. It is the role of phonetics to provide a mapping from these phonological elements to continuous acoustic parameters.

- There are 4 dimensions of intonational performance (Liberman & Pierrehumbert, 1984):
 - tune: intonation contour
 - prominence: local degree of stress or emphasis
 - pitch range: global, or at least phrase-sized, choice of pitch scaling parameters
 - declination: downward trend in pitch across a phrase



- Intonation also requires an interface with semantics and syntax.

11. Pitch variations that affect the meaning (lexical or morphological) of a word are called tone or lexical tone.

- Some languages (e.g., Shona, Chinese) specify the pitch as well as the segmental quality of each vowel.
- In the simplest tone cases (like Shona), there are just two tones: H and L.
- In more complex tonal systems (like Mandarin), there are more tone contrasts, and they cannot be described in terms of single points within a pitch range. The speaker's aim is to produce a pitch movement, which we call a contour tone.

tone 1: high level	˥	ma 'mother'
tone 2: high rising	˧˥	ma 'hemp'
tone 3: low falling rising	˨˩˦	ma 'horse'
tone 4: high falling	˥˧	ma 'scold'

- Tones may be changed due to the influence of adjacent tones: tone sandhi.
- Tones are subject to declination.

12. A lexical pitch accent is emphasis that results just from pitch, rather than loudness or duration.

- Some languages (e.g., Japanese, also Swedish) use pitch accents rather than stress to mark prominence within a word.
- In Japanese, pitch accents are realized as HL on one vowel in some (most) words. But unlike lexical tone languages (e.g., Yoruba), pitch events are not required on every syllable. And unlike in stress accent languages (e.g., English), a pitch event is not even required *somewhere*.

in Tokyo Japanese:

H	umái
L	* umai
	uma <u>l</u> i

1. If the accent is on the 1st syllable, then the 1st syl. is high and the others are low: H-L, H-L-L, etc.
2. If the accent is on a syllable other than the first, then the 1st syllable is low, following sylls up to and including the accented one are high, and the rest are low: L-H, L-H-L, L-H-H-L, L-H-H, etc.
3. If the word doesn't have an accent, the first syllable is low and everything else is high.

Phonetic realization:

- If the accent is on the first syllable, then the pitch starts high and drops suddenly at the second syllable, then goes down more slowly.
- If the accent is on a syllable other than the first or the last, then the pitch rises gradually until the syllable after the accented syllable, and then goes down suddenly. A native speaker hears the accented vowel as higher than the rest, even though the maximum pitch is actually in the next syllable.
- If the word doesn't have an accent, the pitch rises continuously from a low at the start of the word to a high at its end, just like French. About 80% of all Japanese words belong to this class.

Pitch tracking & pitch in Praat

1. What is pitch?
 - the auditory sensation of tonal height
 - In speech, this sensation reflects the periodicity of the speech signal.
 - The vocal folds vibrate during voiced sounds. This vibration creates regular (periodic) fluctuations of air pressure that impinge upon a hearer's eardrum.
 - The rate of vibration (calculated as the number of vibrations per second = Hz) is the fundamental frequency (f_0). The percept of fundamental frequency is pitch.
 - $F_0 = 1 / T$
 - higher $f_0 \rightarrow$ higher pitch
 - Speakers control vocal fold tension during phonation to influence pitch.
 - Note this means that pitch only exists for voiced speech sounds.
2. Computer pitch tracking (f_0 tracking)
 - Pitch-tracking algorithms: autocorrelation → see §2.3.1 (Johnson) for more details
 - The algorithm examines values in an analysis window (that is generally quite a bit longer than a pitch period). The window is moved along, sample by sample, and the values in the window at each point are compared to (correlated with) the values at the window's starting position. Only when the window begins at an equivalent point in a following pitch period will the correlation be high. In this way, repeating patterns (i.e., pitch periods) can be found.
 - f_0 contours do not directly represent either pitch or even signal properties
3. Using Praat to measure F_0
 - To turn on pitch analysis, select "Show pitch" in the "Pitch" menu in the Sound window. All of the default pitch settings should be ok.
 - You will probably also want to turn off formant tracking (uncheck "Show formants" in the "Formant" menu) to see the pitch track better.
 - To start a log for your measurements, go to "Log settings..." in the "Query" menu on the "Sound" window.
 - You will have to change one of your log files (1 or 2)—you can pick.
 - Choose a location and file name for "Log file" I recommend something like:
C:\Documents and Settings\All Users\Desktop\F0 Log.txt
 - In "Log format:", type the following:
't1:4' 'tab\$' 'f0:2' (yes, type the single quotes)
This will give you the time (t_1) and fundamental frequency (f_0) at the cursor point. Click "ok".
 - Now you can record F_0 by simply choosing the relevant point in the pitch track and hitting F12 (for log 1) or Shift-F12 (for log 2).
 - Praat allows you to automatically move your cursor to the F_0 max or min in a selection. After highlighting the relevant vowel, hit Ctrl-L to move to the F_0 minimum (L for Lo) or hit Ctrl-H to move to the F_0 maximum (H for Hi).

4. Pitch tracking errors

- pitch halving: two periods are treated as one
- pitch doubling: one period is treated as two
- failure to detect periodicity

5. Pitch tracking parameters in Praat

Sometimes, adjusting the parameters of the pitch tracking algorithm will help to reduce these errors.

- *pitch range* (default 75-500 Hz)
 - Minimum sets the length of the autocorrelation analysis window (3 periods at the minimum f0), so no pitch can be detected below the minimum. Candidate f0 points above the maximum are not considered.
- *optimize for*: intonation (AC) – to use an autocorrelation method of pitch tracking

post-processing parameters (in Advanced pitch settings...):

- *silence threshold* – amplitude threshold for speech
- *voicing threshold* – affects the algorithm's decision about whether voicing is present; higher values yield more voiceless decisions (i.e., more intervals without pitch) [lower this if voiced segments aren't yielding a pitch track]
- *octave cost* – there may be peaks in the autocorrelation function at multiples of the fundamental (if the analysis window is long enough); higher values favor higher f0s [raise this if you are getting pitch-halving]
- *octave-jump cost* – affects the algorithm's decision about whether a jump in f0 is reasonable; larger values disfavor abrupt changes in f0 [increase this if you are getting pitch-doubling; decrease it if you are failing to track actual rapid changes in f0]
- *voiced/unvoiced cost* – [increase this value to decrease the number of voiced/unvoiced transitions]

6. Segmental effects – microprosody

- Even with a good pitch track, it can be difficult to decide what is ‘linguistically significant’.
- Segments cause perturbations in the f0 contour
 - edge effects – transitions from voiceless to voiced can often be accompanied by disruptions of regular periodicity
 - aerodynamic effects – the rate of vocal fold vibration is affected by the rate of airflow through the glottis, which is in turn affected by articulations related to segmental distinctions
- obstruents cause f0 to lower at closure and raise at release; lowering and raising are greater with voiced obstruents
- sonorants appear to lower f0 through their duration
- high vowels have higher intrinsic f0
- of course, voiceless segments don’t have an f0